PARTNERSHIPS & COLLABORATION

TAPPING INTO COLLECTIVE INTELLIGENCE: MAKING A BIG PROJECT WORK AT A SMALL INSTITUTION

Rebecca A. Morin
California Academy of Sciences
55 Music Concourse Drive
San Francisco, CA 94118
rmorin@calacademy.org

Abstract:
The model of “collective intelligence” is one of the most important concepts for any librarian or information professional to grasp in the 21st century. In times of budget reductions and rapid technological advancement, the best research and most timely innovations cannot be accomplished in a vacuum. This paper outlines the concept of collective intelligence, its importance in the current academic landscape, and the experiences of the California Academy of Sciences in executing a major collaborative grant-funded project.

Keywords: Collective intelligence, innovation, academic libraries, collaboration.

One of the major challenges faced by librarians in small institutions is the burden of writing grants and executing meaningful projects in work environments short on staff and space, and high in demands and expectations. When so many resources are dedicated to everyday tasks, how does a librarian find the time and bandwidth to go above and beyond? One answer is through skillfully employing Collective Intelligence.

The way many think of collective intelligence is in the form of “consensus decision making” in the animal kingdom, which can take the form of the behavior of flocking birds, which do not appear to have a leader, but rather function as something of a self-organized system with individuals making adjustments and decisions in concert with their neighbors. The construction of nests in social insects such as bees, termites, or ants is another example of collective intelligence. There is also the means by which bees select the best fields of flowers from which to harvest nectar, and how they work their way systematically through the chosen area. These are situations in which huge numbers of very simple components (birds, bees, ants) can cope with extremely complex tasks and problems as a group, far above their capacity as individuals.

Collective Intelligence can be defined, simply, as the ability of a group to solve more problems than its individual members. Thomas W. Malone is the founding director of the MIT Center for Collective Intelligence. He says that “As all the people and computers on our planet get more and more closely connected, it's becoming increasingly useful to think of all the people and computers on the planet as a kind of global brain.” The idea of working in the environment of the “global brain” is nothing new to most librarians, particularly the body of professionals in IAMSLIC; groups such as this exist because information professionals value collaboration and free exchange of ideas. However, most librarians have been part of a TERRIBLE collaboration. Librarians love to form groups and clubs and subcommittees and consortia, and sometimes it’s an utter disaster. Sometimes it is so awful that it’s easy to think about how
much simpler it is to confront challenges alone and insulate oneself from the rough waters of collaboration.

Going alone is quite simply not a recipe for success. One will succeed in some areas, but the Lone Wolf strategy sets one up to fall short of major accomplishments. The example of the tiger is illustrative. Tigers are fairly solitary and they hunt alone. This makes sense as a stealth predator. Sneaking up on something is hard enough when alone, and nearly impossible as a group of giant, deadly cats hunting together. After a kill, a group of cats would then have to share the kill. Simply put, it is not economical for tigers to work together in this way, because FLYING UNDER THE RADAR is the defining element of this cat’s existence. To bring the example back to the Library and Information Management arena, one can consider the boss or coworker whose entire approach to life is to keep his or her head down and not draw any attention. Not only are such coworkers difficult to work with, but theirs also are often the first positions trimmed when budgets are tight.

There are downsides to the other extreme as well, illustrated via another zoological example, that of the Carolina Parakeet. Now extinct, it was the only parrot native to North America. Behavior exhibited in flocking, as mentioned earlier, is an example of collective intelligence and consensus decision-making. However, just as merely collaborating with peers will not solve all professional problems, the Carolina Parakeet’s flock behavior helped to lead to its downfall. These birds were known to flock together, to sleep in large groups in the hollows of trees, and if a member of the flock became injured, large numbers would crowd around the wounded party. These behaviors made it incredibly easy for farmers, who disliked the toll birds were taking on their crops, to net and destroy large populations at once. One could destroy a sleeping flock, or injure one bird and then net the dozens of others who would flock to its aid. The extinction of this species is directly related to a set of cooperative behaviors that failed to work advantageously under changing circumstances. The challenge all information professionals face is finding the correct balance and hitting the moving target. A librarian cannot be a tiger anymore, but must also avoid being a Carolina Parakeet.

When under pressure, people all tend to think the same way and deliver the same answers. It is not that those who answer differently are more intelligent or somehow better prepared for life, they are just approach this particular set of circumstances a bit differently. When librarians design projects, plan papers, and write software, the natural tendency is to target like audiences—these audiences generally being other librarians, primary patrons, and others encountered in the course of daily business. This is one of the reasons fruitful collaborations can be so difficult— if one’s work is a reflection of oneself, values, and what individuals find most important. It is easy to decide that this is easiest, and the correct way to operate. In other words, people may react like Carolina Parakeets much of the time, but people also like being tigers.

Any dexterity this author may claim in navigating the Collective Intelligence landscape can be discuss in the context of a large, highly collaborative project operating out of the California Academy of Sciences. Connecting Content: A Collaboration to Link Field Notes to Specimens and Published Literature is a National Leadership Grant project funded by the Institute of Museum and Library Services (IMLS), for which the author serves as Principal Investigator. This project’s 21st century goals are well illustrated via a lovely 16th century image.
This image is an engraving showing the Natural History collection of Ferrante Imperato, an apothecary from Naples, who published this image with the catalog of his collection in 1599. The goal of Connecting Content is to bring everything in this room here conceptually together in a digital environment, under open access principles. One can think about today’s natural history museum as, in many ways, unchanged from what is shown in this illustration. There are specimens, of course, and related to those specimens, there are published materials, such as the books on the far right of the engraving. The young man with the right explaining the collections to the assembled is said to Imperato’s son, who helped him write his observations of the collections—a scholar who composed significant pre-publication materials. All of these elements are intrinsically linked to each other, and there is special knowledge held not just in each element, but also in the relationships between these objects.

Through this project, scholarly institutions are harnessing unique content—specimens from biological collections and field notes from our archives, and digitizing them. All of this content lives in different places in a museum, and increasingly, more of this content has an online representation (or at least significant metadata). Project participants are creating or harnessing this content and connecting it to related published materials that are already digitized and served online via the Biodiversity Heritage Library. The Connecting Content project grew out the Biodiversity Heritage Library, a consortium of natural history and botanical libraries that cooperate to digitize and make accessible the legacy literature of biodiversity held in their collections and to make that literature available for open access and responsible use as a part of a global “biodiversity commons.”
The Connecting Content project operates out of a small library, one that cannot participate in such projects without finding other like-minded individuals and building strong partnerships. As mentioned previously, partnerships are not easy or self-sustaining. It is easy to join a consortium and not do any work. It’s also easy to join one and find yourself in over your head. There are six institutions participating in the Connecting Content project, and they all members of and contributors to the Biodiversity Heritage Library. This gives the institutions a close relationship, but they also have access to BHL resources and expertise— for example, that is why the project serves content via the BHL portal. It makes sense for BHL, it makes sense for Connecting Content, and the projects do not squander resources building a new delivery system.

As part of Connecting Content, the participating institutions have created pilot projects, themed collections of archival and specimen material, to digitize and contribute to the project as proof of concept. Part of what makes this collaboration interesting and useful is that each of these organizations possesses varying built-in resources, workflows, and procedures. Partners learn from each other and create projects that are scalable and extensible. This does not mean that the project is simple, but some problems have been solved via robust collaboration. To the scientific staff of a natural history museum, the idea that there are direct, multi-way relationships between specimens, field notes, and published literature makes perfect sense. In fact, the project was inspired in part by the fact that researchers regularly visit institutions to begin work in one department and end up being traded off between different staff members who serve as custodians of different pieces of the collections puzzle. For example, someone may come in through the herpetology department looking at specimens. It is very common for that person to consult the published literature as well, a new wrinkle that probably involves working with library staff. At some point in the process, the researcher may have a question about something written in the published papers or discovered in relation to a specimen, which may then require a look at the original field notes documenting the collection activity itself. Those field notes may live in the Archives. They might be in the Herpetology department. They might be the field notes of someone who collected in more than one discipline, in which case those herpetology notes might be in the ornithology department. At least three different members of the scientific staff have become involved in the answering of this question. Connecting Content is, on its face, an attempt to streamline this process and co-locate information about the same item in a manner that is easy to find.

The project partners thought this would be fairly simple, at least on a conceptual level. A great deal of preliminary work with collections managers and curators took place before any digitization work took place, but some issues were overlooked, stumbling blocks that that in hindsight should have been obvious. The project partners failed to exploit collective intelligence properly. A big problem was how to use a display mechanism and service designed to serve published books and journals to deliver archival content, a challenge the partners dubbed “fitting the archives peg into a book-shaped hole,” and it was never satisfactorily resolved. The emphasis on creating new digital content (digitizing field notes and specimens) required so much energy and that it took a programmer not working on this grant to call attention to problems related to item display and provision of metadata. It took another archivist to raise conceptual questions about original order and artificial collections.

Considering a MODS record created for this project, one can see it is not a robust chunk of metadata. An item from an archival collection is not really meant to stand alone as an item the way a monograph does. Archival materials are meant to be taken in context with the other items in the collections, and that is not really possible using the BHL interface and book viewer.

Other problems arose around content creation. Digitizing field notes is fairly straightforward, as the task can be accomplished using a flatbed scanner, a camera stand for overhead photographs, or a book scanner. It may not be cheap or fast, but it is inherently doable. But, several of the grant partners, including the California Academy of Sciences, are digitizing bird specimens as part of their pilot projects. How does one digitize a bird? What visual information do people need to find an image of a bird specimen useful? Each institution developed different methods to capture as much useful information as possible. Since this is a highly collaborative effort, and the idea is that each pilot project could serve as a model for other institutions to try their own projects is key. With institutions of varying sizes and institutional capacities contributing pilot projects, the hope is that this will demonstrate scalability. Success in this project will be to see other institutions learn from Connecting Content’s mistakes and find a description and digitization model in one of our pilot projects that works for them.

So far in the discussion of Connecting Content and learning, the “student” at this point is not the end-user, but rather the grant partners—other librarians, archivists, and scientists. However, these individuals are not the key learners and the target audience for this project.
To return to another zoological metaphor, this is *Cactospiza pallida*, or *Camarhynchus pallidus* (depending which name the reader accepts as valid), a species of finch from the Galapagos archipelago, one of the varieties of finch collected by Charles Darwin during the Beagle voyage. This particular species of finch actually uses tools. In addition to serving as examples of specialization via natural selection, these finches have shown the ability not just to use tools, but to change their tool use on the fly to obtain food from various containers designed to force them to adapt their behaviors. It is important to keep organisms like this finch in mind when considering who is served by a project like *Connecting Content*, and how those patrons are creative, flexible, adaptable, and in many ways, at least one step ahead of the librarians and information professionals spearheading such projects. In fact, the missing piece of many successful collaborations is how to tap into the communities of users in ways that make projects better. IMLS calls this concept “putting the learner at the center” and it is a key piece of the *Connecting Content* puzzle. During the crafting of the *Connecting Content* proposal, an IMLS program officer told the partners that nobody needs another portal. So instead of creating terabytes of content and serving it via a static online portal, the aim for deliverables is to turn to some major user communities for the creation of third-party web applications that harness the created open content. In some ways, the final grant deliverable is a product not designed by the participants in the project. As mentioned earlier, the easiest projects are designed to serve reflections of oneself, and this group decided early that the best way to avoid creating yet another tool or resource destined to rot away unmaintained after the money ran out was to put faith into the hands of the target audiences and see what sort of user generated, creative, non-prescribed, fast, lightweight, inventive tools and services these folks could come up with. An oft-repeated mantra for this project is that it is built upon the concept that is not the job of the content creator (the librarian, archivist, or scientist) to tell the user what to do with the data. These professionals know how they want to use their data, but that is of little use to those outside those communities. As the *Connecting Content* project draws to a close in the Spring of 2014, the partners must confront how to deliver on promises made. Much of the linking promised is based on the ability of users to have the datapoints required for useful links. When working with handwritten, nonstandard materials, the generation of those datapoints is often based in human work, involving transcription (for example). This is another learning experience for the project partners, and a stark reminder that while the community has figured out some things about digitization, data delivery, and automation, there is a long way to go. The low-hanging fruit is running out, and climbing further up the tree is not easy. The project has employed students and interns to work on identifying the scientific names in the field note scans, and tagging the pages in BHL with those names so that the BHL discovery tools can incorporate the field notes into species bibliographies, even though the pages do not have usable OCR text.

It is important to keep in mind that what people want and what we currently have the institutional capacity to deliver are widely different things. There is a quote from an episode of *Doctor Who* from the 1970s: “Answers are easy. It’s asking the right questions which is hard.” This sums up why librarians and information professionals need to not only build robust collaborative relationships with peers, but also break the mold by offering a larger, participatory role to users. One cannot really tap into the collective intelligence of a user community without clearly defining your problems. And the easiest way to define those problems is via users.
Figure 4. Image of Rollo Beck field notes from the Biodiversity Heritage Library (left) with the uncorrected OCR (right). http://biodiversitylibrary.org/item/121726 #page/6/mode/1up (1 October 2013).

References